

Agnostic insurability of model classes

Narayana Santhanam

*Dept. of Electrical Engg,
University of Hawaii at Manoa
e-mail: nsanthan@hawaii.edu*

and

Venkat Anantharam

*Dept of EECS
University of California, Berkeley
e-mail: ananth@eecs.berkeley.edu*

Abstract: Motivated by problems in insurance, our task is to predict finite upper bounds on a future draw from an unknown distribution p over the set of natural numbers, using only past observations generated independently and identically distributed according to p . While p is unknown, it is known to belong to a given collection \mathcal{P} of probability distributions on the natural numbers. The support of the distributions $p \in \mathcal{P}$ may be unbounded, and the prediction occurs for *infinitely* many draws. We are allowed to make observations without predicting upper bounds for some time, but must start and then continue to predict upper bounds after a finite time with probability 1 irrespective of which $p \in \mathcal{P}$ governs the observations.

If it is possible, without knowledge of p and for any prescribed confidence however close to 1, to come up with a sequence of upper bounds that is never violated over an infinite time window with confidence at least as big as prescribed, we say the model class \mathcal{P} is *insurable*. We characterize insurability of any class \mathcal{P} of distributions over natural numbers by a condition on how the neighborhood of distributions $p \in \mathcal{P}$ should behave, one that is both necessary and sufficient.

Keywords: insurance, ℓ_1 topology, non-parametric approaches, prediction, universal compression.

1. Introduction

Insurance is a means of managing risk by transferring a potential sequence of losses to an *insurer* for a price paid at the beginning of each period, called the *premium*. The insurer attempts to break even by balancing the possible loss that may be suffered by a few with the guaranteed premiums of many. We aim to study the fundamentals of this problem when the losses

can be unbounded and a precise model for the probability distribution of the aggregate loss in each period either does not exist or is infeasible to get.

A systematic, theoretical, as opposed to empirical, study of insurance goes back to 1903 when Filip Lundberg [1] defined a natural probabilistic setting as part of his thesis. In particular, Lundberg formulated a collective risk problem pooling together the risk of all the insured parties into a single entity, which we call the insured. Typically, studies of insurance derived from the approach of [1] depend on working with specific models for the loss distribution, e.g. compound Poisson models, after which questions of interest in practice, such as for instance the relation between the size of the premiums charged and the probability of the insurer going bankrupt, can be analyzed. A rather comprehensive theory of insurance along these lines has evolved [2, 3] which incorporates several model classes for the distribution of the losses over time other than compound Poisson processes, and which also includes some heavy tailed distribution classes.

We depart from the existing literature in two important respects. The first relates to the practice among insurers to limit payments to a predetermined ceiling, even if the loss suffered by the insured exceeds this ceiling. In both the insurance industry and the legal regulatory framework surrounding it, this is assumed to be common sense. But is it always necessary to impose such ceilings? Moreover, in scenarios such as reinsurance, a ceiling on compensation is not only undesirable, but may also limit the very utility of the business.

The second unconventional aspect of our approach arises from our motivation to deal with several new settings for which some sort of insurance is desirable, but where insurers are hesitant to enter the market. Examples of such settings include insuring against network outages or attacks against future smart grids, where the cascade effect of outages or attacks could be catastrophic. In these settings, it is not clear today what should constitute a reasonable risk model because of the absence of usable information about what might cause the outages or motivate the attacks. We address this issue by working with a *class* of models, i.e. a set of probability laws over loss sequences that adheres to any assumptions the insurer may want to make or any information it may already have. In this paper we will only consider loss models that are independent and identically distributed (i.i.d.) from period to period, so we can equivalently think of a model class as defined in terms of its one dimensional marginals. As an example, we may want to consider the set of all finite moment probability distributions over the nonnegative integers as our class of possible models for the loss distribution in each period. Now, we ask the question: what classes of models are the

ones on which the insurer can learn from observations and set premiums so as to remain solvent? In this paper, we completely answer this question by giving a necessary and sufficient condition that characterizes what classes of models lend themselves to this insurance task.

Formally, we adopt the collective risk approach, namely, we abstract the problem to include just two agents, the insurer and the insured. Losses incurred by the insured are considered to form a discrete time sequence of random variables, with the sequence of losses denoted by $\{X_i, i \geq 1\}$, and we assume that $X_i \in \mathbb{N}$ for all $i \geq 1$, where \mathbb{N} denotes the set of natural numbers, $\{0, 1, 2, \dots\}$. A *model class* \mathcal{P}^∞ is a collection of measures on infinite length loss sequences, and is to be thought of as the set of all potential probability laws governing the loss sequence. Each element of \mathcal{P}^∞ is a *model* for the sequence of losses. Any prior knowledge on the structure of the problem is accounted for in the definition of \mathcal{P}^∞ . We focus on measures corresponding to *i.i.d.* samples, i.e. each member of \mathcal{P}^∞ is a product distribution. We denote by \mathcal{P} the set of distributions on \mathbb{N} obtained as one dimensional marginals of \mathcal{P}^∞ . Since there is no risk of confusion, we will also refer to the distributions in \mathcal{P} as models and to \mathcal{P} as the model class.

The actual model in \mathcal{P} governing the law of the loss in each period remains unknown to the insurer. We assume no ceiling on the loss, and require the insurer to compensate the insured in full for the loss in each period at the end of that period. The insurer is assumed to start with some initial capital $\Pi_0 \in \mathbb{R}^+$, a nonnegative real number. The insurer then sets a sequence of premiums based on the past losses—at time i , the insurer collects a premium $\Pi(X_1^{i-1})$ at the beginning of the period, and pays out to compensate for the loss X_i at the end of the period. If the built up capital till step i (including $\Pi(X_1^{i-1})$, and after having paid out all past losses) is less than X_i , the insurer is said to be *bankrupted*. Given a class \mathcal{P}^∞ of loss models, we ask if for every prescribed upper bound $\eta > 0$ on the probability of bankruptcy, the insurer can set (finite) premiums at every time step based only on the loss sequence observed thus far and with no further knowledge of which law $p \in \mathcal{P}^\infty$ governs the loss sequence, such that the insurer remains solvent with probability bigger than $1 - \eta$ under p irrespective of which $p \in \mathcal{P}^\infty$ is in effect. If the probability of the insurer ever going bankrupt over an infinite time window can be made arbitrarily small in this sense, the class of *i.i.d.* loss measures \mathcal{P}^∞ is said to be *insurable*.

A couple of clarifications are in order here. First, to make the problem non-trivial, we allow the insurer to observe the loss sequence for some arbitrary finite length of time without having to provide insurance. We require that the insurer has to eventually provide insurance with probability 1 no matter

which $p \in \mathcal{P}^\infty$ is in effect, and cannot quit providing insurance once it has entered into the insurance contract with the insured. Premiums set before the entry time can be thought of as being 0 and the question of bankruptcy only arises after the insurer has entered into the contract. Secondly, we do not concern ourselves with incentive compatibility issues on the part of the insured and assume that the insured will accept the contract once the insurer has entered, agreeing to pay the premiums as set by the insurer.

It turns out that the fact that the capital available to the insurer at any time is built up from past premiums does not play any role in whether a model class is insurable or not. In fact, the problem is basically one of finding a sequence of finite upper bounds $\Phi(X_1^{i-1})$ on the loss X_i for all $i \geq 1$. We refer to the sequence $\{\Phi(X_1^{i-1}), i \geq 1\}$ as the *loss dominating sequence* and call $\Phi(X_1^{i-1})$ the *loss-dominant* at step i . The notion of insurability of a model class \mathcal{P} comes down to whether for each $\eta > 0$ there is a way of choosing the loss dominants in such a way that the probability of the loss X_i ever exceeding the loss dominant $\Phi(X_1^{i-1})$ is smaller than η irrespective of which model p in the model class \mathcal{P}^∞ is in effect. Here again we allow some initial finite number of periods for which the loss dominant can be set to ∞ , but it must become finite with probability 1 under each $p \in \mathcal{P}^\infty$ and stay finite from that point onwards.

Theoretically, the flexibility we have permitted regarding when to start proposing finite loss dominants places the insurance problem formulated above in a class of problems that can be said to admit *useful* pointwise convergent estimates. Roughly speaking, the insurance problem can be thought of as one of requiring estimating all the percentiles of an unknown distribution from \mathcal{P} , using only *i.i.d.* draws generated from it. However, as the sample size increases, the estimate of any given percentile need not converge to the true value (according to some predefined metric) uniformly over the entire class \mathcal{P} . Even if the estimate converges only pointwise over the class, it is useful if, for any given finite sample size, we could also say whether the estimate is doing well or not relative to the true model even though we don't know what the true model is. This is the case for the notion of insurability of a model class that we have introduced above. More generally, when dealing with large alphabets or high dimensions, it is sometimes too restrictive to require estimates or algorithms to converge to the true values uniformly over the model class as the sample size increases to infinity. Instead, this notion of useful pointwise convergence allows us to consider broader model classes from a practical perspective. For other kinds of such pointwise estimation, particularly in relation to information theoretic quantities, see [4].

For a model class to be insurable, roughly speaking, close distributions

must have comparable percentiles. Distributions in the model class that, in every neighborhood, have some other distribution with arbitrarily different percentiles are said to be *deceptive*. In Section 2, we define what it means for distributions to be close, and what it means for distributions to have comparable percentiles. In Section 3, we provide several examples of insurable and non-insurable model classes. Our main result is Theorem 1, which states that that \mathcal{P}^∞ is insurable iff it has no deceptive distributions. We prove this theorem in Sections 4 and 5.

2. Precise formulation of the problem and statement of the main result

We model the loss at each time by a random variable taking values in $\mathbb{N} = \{0, 1, \dots\}$. Denote the sequence of losses by X_1, X_2, \dots where $X_i \in \mathbb{N}$. Let \mathbb{N}^* be the set of all finite length sequences from \mathbb{N} , including the empty sequence. We will write x^n for the sequence x_1, \dots, x_n . Where it appears, x^0 denotes the empty sequence. A loss distribution is a probability distribution on \mathbb{N} . Let \mathcal{P} be a set of loss distributions. \mathcal{P}^∞ is the collection of *i.i.d.* measures over infinite sequences of symbols from \mathbb{N} such that the set of one dimensional marginals over \mathbb{N} they induce is \mathcal{P} .

We write \mathbb{R}^+ for the set of nonnegative real numbers and use $:=$ for equality by definition.

Consider an insurer with an *initial capital* $\Pi_0 \in \mathbb{R}^+$. An *insurance scheme* for \mathcal{P} is comprised of a pair (τ, Π) . Here $\tau : \mathbb{N}^* \mapsto \{0, 1\}$ satisfies $\tau(x_1, \dots, x_n) = 1 \implies \tau(x_1, \dots, x_{n+1}) = 1$ for all x^n and also $p(\sup_n \tau(X^n) = 1) = 1$ for all $p \in \mathcal{P}^\infty$. τ should be thought of as defining an *entry time* for the insurer with the property that once the insurer has entered it stays entered and that the insurer enters with probability 1 irrespective of which $p \in \mathcal{P}^\infty$ is in effect. Here we say the insurer enters after seeing the sequence $x^n \in \mathbb{N}^*$ (possibly the empty sequence) if $\tau(x^n) = 1$. The other ingredient of an insurance scheme is the *premium setting scheme* $\Pi : \mathbb{N}^* \rightarrow \mathbb{R}^+$, satisfying $\Pi(x^n) = 0$ if $\tau(x^n) = 0$, with $\Pi(x^n)$ being interpreted as the premium demanded by the insurer from the insured after the loss sequence $x^n \in \mathbb{N}^*$ is observed.

Let $1(\cdot)$ denote the indicator function of its argument. The event that the insurer goes bankrupt is the event that

$$\Pi_0 + \sum_{i=1}^n (\Pi(X^{i-1}) - X_i) 1(\tau(X^{i-1}) = 1) < 0 \text{ for some } n \geq 1.$$

In words, this is the event that in some period $n \geq 1$ after the insurer has entered, the loss X_n incurred by the insured exceeds the built up capital of

the insurer, namely the sum of its initial capital and all the premiums it has collected after it has entered (including the currently charged premium $\Pi(X^{n-1})$) less all the losses paid out so far.

Definition 1. A class \mathcal{P}^∞ of laws on loss sequences is called insurable by an insurer with initial capital $\Pi_0 \in \mathbb{R}^+$ if $\forall \eta > 0$, there exists an insurance scheme (τ, Π) such that $\forall p \in \mathcal{P}^\infty$,

$$p((\tau, \Pi) \text{ goes bankrupt}) < \eta .$$

We should remark that despite the apparent role of the initial capital of the insurer in this definition, it plays no role from a mathematical point of view. To see this note first that if a model class \mathcal{P}^∞ is insurable by an insurer with capital Π_0 it is clearly insurable by all insurers with initial capital at least Π_0 , since such an insurer can use the same entry time and premium setting scheme as the insurer with initial capital Π_0 . On the other hand, an insurer with initial capital less than Π_0 can use the same entry time as an insurer with initial capital Π_0 and simply charge an additional premium at the time of entry which in effect builds up its initial capital to Π_0 , and then proceed with the same premium setting scheme as that used by the insurer with initial capital Π_0 . This feature is an artifact of the complete flexibility we give the insurer in setting premiums; for more on this see the concluding remarks in Section 6.

As indicated in the introductory Section 1, we will first show that whether a model class of loss distributions is insurable is equivalent to whether we can find suitable loss domination sequences for the sequence of losses. We next make this connection and the associated terminology precise.

A *loss domination scheme* for \mathcal{P} is a mapping $\Phi : \mathbb{N}^* \mapsto \mathbb{R}^+ \cup \{\infty\}$, where for $x^n \in \mathbb{N}^*$, we interpret $\Phi(x^n)$ as an estimated upper bound on x_{n+1} . We call $\{\Phi(X^{i-1}), i \geq 1\}$ the loss-domination sequence and $\Phi(X^{i-1})$ the loss-dominant at step i . We require for all $x^n \in \mathbb{N}^*$ that

$$\Phi(x_1, \dots, x_n) < \infty \implies \Phi(x_1, \dots, x_{n+1}) < \infty$$

and also that for all $p \in \mathcal{P}^\infty$,

$$p(\inf_{n \geq 1} \Phi(X^n) < \infty) = 1.$$

We think of $\Phi(x^n) = \infty$ as saying that the scheme has not yet committed to proposing finite loss dominants after having seen the sequence x^n , while if $\Phi(x^n) < \infty$ it has. Once the scheme commits to proposing finite loss

dominants it has to continue to propose finite loss dominants from that point onwards. Further, with probability 1 under every $p \in \mathcal{P}^\infty$, the scheme has to eventually start proposing finite loss dominants. Given our motivation from the insurance problem, we will say the loss domination scheme Φ goes *bankrupt* if $\Phi(X^{n-1}) < X_n$ for some $n \geq 1$.

The connection between the insurance problem and the problem of selecting loss dominants can now be made precise as follows.

Observation 1. Let \mathcal{P}^∞ be a model class and $\eta > 0$. Let $\Pi_0 \in \mathbb{R}^+$. An insurer with initial capital Π_0 can find an insurance scheme (τ, Π) such that the probability of remaining solvent is bigger than $1 - \eta$ irrespective of which $p \in \mathcal{P}^\infty$ is in effect if and only if there is a loss domination scheme Φ such that the probability of it going bankrupt is less than η irrespective of which $p \in \mathcal{P}^\infty$ is in effect.

Proof Given an insurance scheme (τ, Π) consider the loss domination scheme Φ that has $\Phi(x^n) := \infty$ iff $\tau(x^n) = 0$ and

$$\Phi(X^{n-1}) := \Pi_0 + \sum_{i=1}^{n-1} (\Pi(X^{i-1}) - X_i) 1(\tau(X^{i-1}) = 1) + \Pi(X^{n-1}),$$

if $\tau(X^n) = 1$. Since τ enters (become equal to 1) with probability 1 under each $p \in \mathcal{P}^\infty$ and stays equal to 1 once it has become 1, Φ becomes finite with probability 1 under each $p \in \mathcal{P}^\infty$ and stays finite once it has become finite. Thus Φ is indeed a loss domination scheme. It is straightforward to check that if the insurance scheme (τ, Π) stays solvent with probability bigger than $1 - \eta$ irrespective of which $p \in \mathcal{P}^\infty$ is in effect then the loss domination scheme Φ becomes bankrupt with probability less than η irrespective of which $p \in \mathcal{P}^\infty$ is in effect.

Conversely, given a loss domination scheme Φ define the insurance scheme (τ, Π) by setting $\tau(x^n) := 0$ iff $\Phi(X^n) = \infty$ (and $\tau(x^n) := 1$ iff $\Phi(x^n) < \infty$) and defining $\Pi(x^n) := 0$ if $\Phi(x^n) = \infty$ and $\Pi(x^n) := \Phi(x^n)$ if $\Phi(x^n) < \infty$. One sees that τ as defined becomes 1 with probability 1 under each $p \in \mathcal{P}^\infty$ and stays equal to 1 once it becomes 1. Further, the premiums set at each time are finite and equal to 0 till the entry time. Thus (τ, Π) as defined is indeed an insurance scheme. It is straightforward to check if Φ becomes bankrupt with probability less than η irrespective of which $p \in \mathcal{P}^\infty$ is in effect, then (τ, Π) stays solvent with probability bigger than $1 - \eta$ irrespective of which $p \in \mathcal{P}^\infty$ is in effect. Hence the above observation. \square

We may therefore conclude that a model class \mathcal{P}^∞ is insurable iff for all $\eta > 0$ there is a loss domination scheme Φ such that the probability of going

bankrupt under Φ is less than η irrespective of which $p \in \mathcal{P}^\infty$ is in effect. In the rest of the paper we will therefore focus mainly on whether the model class \mathcal{P}^∞ is such that for every $\eta > 0$ a loss domination sequence Φ exists with its probability of bankruptcy being less than η irrespective of which model in the model class governs the sequence of losses.

In Theorem 1, we provide a condition on \mathcal{P} that is both necessary and sufficient for insurability.

2.1. Close distributions

Insurability of \mathcal{P}^∞ depends on the neighborhoods of the probability distributions among its one dimensional marginals \mathcal{P} . The relevant “distance” between distributions in \mathcal{P} that decides the neighborhoods is

$$\mathcal{J}(p, q) := D\left(p \parallel \frac{p+q}{2}\right) + D\left(q \parallel \frac{p+q}{2}\right).$$

Here $D(p \parallel q)$ denotes the relative entropy of p with respect to q , where p and q are probability distributions on \mathbb{N} , defined by

$$D(p \parallel q) := \sum_{y \in \mathbb{N}} p(y) \log \frac{p(y)}{q(y)}.$$

The logarithm is assumed to be taken to base 2 (we use \ln for the logarithm to the natural base).

2.2. Cumulative distribution function

Since we would like to discuss percentiles, it is convenient to use a non-standard definition for the cumulative distribution function of a probability distribution on \mathbb{N} .

For our purposes, the cumulative distribution function of any probability distribution p on \mathbb{N} is a function from $\mathbb{R}^+ \cup \{\infty\} \rightarrow [0, 1]$, and will be denoted by F_p . We obtain F_p by first defining F_p on points in the support of p . We define F_p for all other nonnegative real numbers by linearly interpolating between the values in the support of p . Finally, $F_p(\infty) := 1$.

Let $F_p^{-1} : [0, 1] \mapsto \mathbb{R}^+ \cup \{\infty\}$ denote the inverse function of F_p . Then $F_p^{-1}(x) = 0$ for all $0 \leq x < F_p(0)$. If p has infinite support then $F_p^{-1}(1) = \infty$, else $F_p^{-1}(1)$ is the smallest natural number y such that $F_p(y) = 1$.

Two simple and useful observations can now be made. Consider a probability distribution p with support $\mathcal{A} \subset \mathbb{N}$. For $\delta > 0$, let (T for tail)

$$T_{p,\delta} := \{y \in \mathcal{A} : y \geq F^{-1}(1 - \delta)\},$$

and let (H for head)

$$H_{p,\delta} := \{y \in \mathcal{A} : y \leq 2F^{-1}(1 - \delta/2)\}.$$

It is easy to see that

$$p(T_{p,\delta}) > \delta \text{ and } p(H_{p,\delta}) > 1 - \delta.$$

Suppose that for some $\delta > 0$ we have $F_p^{-1}(1 - \delta) > 0$ and the loss-dominant at the beginning of period $i \geq 1$ happens to be set to $F_p^{-1}(1 - \delta)$, then the probability under p of the loss in period i exceeding the loss-dominant is bigger than δ . If the loss-dominant at the beginning of period i happens to be set to $2F_p^{-1}(1 - \delta/2)$, then the probability that the loss in period i exceeds the loss-dominant is less than δ . We will use these observations in the proofs to follow.

2.3. Necessary and sufficient conditions for insurability

Existence of close distributions with very different quantiles is what kills insurability. A loss domination scheme could be “deceived” by some process $p \in \mathcal{P}^\infty$ into setting low loss-dominants, while a close enough distribution hits the scheme with too high a loss. The conditions for insurability of \mathcal{P}^∞ are phrased in terms of the set of its one dimensional marginals, \mathcal{P} .

Formally, a probability distribution p in \mathcal{P} is called *deceptive* if $\forall \epsilon > 0$, $\exists \delta > 0$ such that that no matter what $f(\delta) \in \mathbb{R}^+$ is chosen, \exists a (bad) distribution $q \in \mathcal{P}$ such that

$$\mathcal{J}(p, q) < \epsilon$$

and

$$F_q^{-1}(1 - \delta) > f(\delta).$$

In the above definition, $f(\delta)$ is simply an arbitrary nonnegative real number. However, it is useful to think of this number as the evaluation of a function $f : (0, 1) \rightarrow \mathbb{R}$ at δ . Equivalently, a distribution p in \mathcal{P} is not *deceptive* if $\exists \epsilon_p > 0$, such that $\forall \delta > 0$, $\exists f(\delta) \in \mathbb{R}$, such that all distributions $q \in \mathcal{P}$ with

$$\mathcal{J}(p, q) < \epsilon$$

satisfy

$$F_q^{-1}(1 - \delta) \leq f(\delta).$$

Our main theorem is the following, which we prove in Sections 4 and 5.

Theorem 1. \mathcal{P}^∞ is insurable, iff no $p \in \mathcal{P}$ is deceptive. \square

3. Examples

Consider \mathcal{U} , the collection of all uniform distributions over a finite contiguous support of the form $\{m, \dots, M\}$, with $m \leq M$ being arbitrary nonnegative integers. Let the losses come as *i.i.d.* samples from one of the distributions in \mathcal{U} —call the resulting model class \mathcal{U}^∞ .

Example 1. \mathcal{U}^∞ is insurable.

Proof If the threshold probability of ruin is η , choose the loss-domination scheme Φ as follows. For all sequences x^n with $n \leq \log \frac{1}{\eta} + 1$ set $\Phi(x^n) = \infty$. For all sequences x^n with $n > \log \frac{1}{\eta} + 1$, the loss-dominant $\Phi(x^n)$ is set to be twice the largest loss observed thus far. It is easy to see that this scheme is bankrupted with probability less than η irrespective of which $p \in \mathcal{U}^\infty$ is in effect. \square

Consider the set \mathcal{N}^∞ of all *i.i.d.* processes such that the one dimensional marginals have finite moment. Namely, $\forall p \in \mathcal{N}^\infty, \mathbb{E}_p X_1 < \infty$.

Example 2. \mathcal{N}^∞ is not insurable.

Proof Note that the loss process that puts probability 1 on the all zero sequence exists in \mathcal{N}^∞ , since it corresponds to the one dimensional marginal loss distribution that produces loss 0 in each period. Since every loss domination scheme enters with probability 1 no matter which $p \in \mathcal{N}^\infty$ is in force, every loss domination scheme must enter after seeing some finite number of zeros. Fix any loss domination scheme Φ . Suppose the scheme starts to set finite dominants after seeing N losses of size 0. To show that \mathcal{N}^∞ is not insurable, we show that $\exists \eta > 0$ and $\exists p \in \mathcal{N}^\infty$ such that

$$p(\Phi \text{ goes bankrupt}) \geq \eta.$$

Fix $\delta = 1 - \eta$. Let ϵ be small enough that

$$(1 - \epsilon)^N > 1 - \delta/2,$$

and let M be a number large enough that

$$(1 - \epsilon)^M < \delta/2.$$

Note that since $1 - \delta/2 \geq \delta/2$, we have $N < M$. Let L be greater than any of loss-dominants set by Φ for the sequences $0^N, 0^{N+1}, \dots, 0^M$. Let $p \in \mathcal{N}^\infty$ satisfy, for all i ,

$$p(X_i) = \begin{cases} 1 - \epsilon & \text{if } X_i = 0 \\ \epsilon & \text{if } X_i = L. \end{cases}$$

For the i.i.d. loss process having the law p , the insurer is bankrupted on all sequences that contain loss L in between the N -th and M -th steps. These sequences, $0^N L, 0^{N+1} L, \dots, 0^{M-1} L$, have respective probabilities (under p)

$$(1 - \epsilon)^N \epsilon, (1 - \epsilon)^{N+1} \epsilon, \dots, (1 - \epsilon)^{M-1} \epsilon,$$

and they also form a prefix free set. Therefore, summing up the geometric series and using the assumptions on ϵ above,

$$p(\Phi \text{ is bankrupted}) \geq (1 - \epsilon)^N - (1 - \epsilon)^M \geq 1 - \delta/2 - \delta/2 = \eta. \quad \square$$

One can actually directly verify that every distribution in \mathcal{N}^∞ is deceptive.

Consider the collection of all i.i.d. loss distributions with monotone one dimensional marginals. A monotone probability distribution p on \mathbb{N} is one that satisfies $p(y + 1) \leq p(y)$ for all $y \in \mathbb{N}$. Let \mathcal{M}^∞ be the set of all i.i.d. loss processes, with one dimensional marginal distribution from \mathcal{M} , the collection of all monotone probability distributions over \mathbb{N} .

Again, it is easily shown that every distribution in \mathcal{M} is deceptive. It follows from Theorem 1 that

Example 3. \mathcal{M}^∞ is not insurable. \square

4. Necessary condition for insurability

In this section we prove one direction of Theorem 1, as stated next.

Theorem 2. If \mathcal{P}^∞ is insurable, then no $p \in \mathcal{P}$ is deceptive.

Proof To keep notation simple, we will denote by p (or q) both a measure in \mathcal{P}^∞ as well as the corresponding one dimensional marginal distribution, which is a member of \mathcal{P} . The context will clarify which of the two is meant. We prove the contrapositive of the theorem: if some $p \in \mathcal{P}$ is deceptive, then \mathcal{P}^∞ is not insurable.

Pick $0 \leq \alpha < h^{-1}(\frac{1}{2})$ where $h(x)$ is the binary entropy function defined for $0 \leq x \leq 1$ by

$$h(x) := -x \log x - (1 - x) \log(1 - x),$$

where the logarithm is to base 2. Fix $0 < \eta < (1 - 2h(\alpha))(1 - \frac{1}{e})$, the bounds chosen in order to satisfy the technical requirements of the proof of Lemma 4. Suppose $p \in \mathcal{P}$ is deceptive. We prove that \mathcal{P}^∞ is not insurable by finding for each loss domination scheme Φ , a probability distribution $q \in \mathcal{P}$ such that

$$q(\Phi \text{ goes bankrupt}) \geq \eta.$$

So, let Φ be any loss domination scheme. Recall that Φ enters on p with probability 1, in the sense that the loss dominants set by Φ will eventually become finite with probability 1 under p . For all $n \geq 1$, let

$$R_n := \{x^n : \Phi(x^n) < \infty\}$$

be the set of sequences of length n on which Φ has entered and let $N \geq 1$ be a number such that

$$p(R_N) > 1 - \alpha/2. \quad (1)$$

For any sequence x^n , let $A(x^n)$ be the set of symbols that appear in it. Recall that the head of the distribution p , $H_{p,\gamma}$, was defined in Section 2.2 to be the set $\{y \in \mathcal{A} : y \leq 2F_p^{-1}(1 - \gamma/2)\}$, where \mathcal{A} is the support of p . Further, define for all $\gamma > 0$

$$R_{p,\gamma,n} := \{x^n \in R_n : A(x^n) \subseteq H_{p,\gamma}\}.$$

Set $\epsilon = \frac{1}{N^2}$. Since p is deceptive, there exists $\delta > 0$ such that for all $f(\delta) \in \mathbb{R}$, there exists a distribution $q \in \mathcal{P}$ satisfying both

$$\mathcal{J}(p, q) < \epsilon = \frac{1}{N^2} \text{ and } F_q^{-1}(1 - \delta) > f(\delta). \quad (2)$$

While the number $f(\delta)$ can be arbitrary above, we focus on a specific number dependent only on Φ . To define this number, first pick $k \geq 2$ large enough that

$$(1 - \delta^k)^{N+1/\delta} \geq 1 - \alpha/2. \quad (3)$$

Note that the limit of the left side above as $k \rightarrow \infty$ is 1, so there is always some choice k that works. Now, for all $0 < \delta' < 1$, let

$$f(\delta') := \max_{\substack{x^i \in R_{p,\delta^k,i} \\ N \leq i \leq N + \lceil \frac{1}{\delta'} \rceil}} \Phi(x^i).$$

A word about this parameter k , since it is not immediately apparent why this should be defined. We will effectively ignore the δ^k tail of the distribution

p , and focus only on strings in $R_{p,\delta^k,i}$, $N \leq i \leq N + \frac{1}{\delta}$. The advantage of doing so is technical—we will be able to handle p and q as though they were distributions with finite span. This is crucial, since we want to have the maximum over a finite set to ensure $f(\delta') < \infty$. Furthermore, note that for $N \leq i < N + \frac{1}{\delta}$, $p(R_{p,\delta^k,i}) \geq 1 - \alpha$ from (1) and (3).

Let $q \in \mathcal{P}$ satisfy (2) with $f(\delta)$ as defined above. Applying Lemma 4 to distributions over length- n sequences induced by the measures $p, q \in \mathcal{P}^\infty$ corresponding to the distributions above,

$$q(R_{p,\delta^k,N}) \geq 1 - \frac{2}{N} - 2h(\alpha),$$

namely, Φ has entered with probability (under q) at least $1 - \frac{2}{N} - 2h(\alpha)$ for length N sequences. Since the insurer cannot quit once it has entered, the scheme has entered with probability (under q) at least $1 - \frac{2}{N} - 2h(\alpha)$ for all n length sequences where $n \geq N$. Namely for all $n \geq N$,

$$q(R_{p,\delta^k,n}) \geq 1 - \frac{2}{N} - 2h(\alpha).$$

For convenience, let $M = \lceil \frac{1}{\delta} \rceil$. Let the distribution q be in force. We have set things up so that Φ is bankrupted whenever any element in the δ -tail of q follows any sequence in $R_{p,\delta^k,i}$, where $N \leq i \leq N + M - 1$. To see this, note that

$$F_q^{-1}(1 - \delta) \geq f(\delta) = \max_{\substack{X^i \in R_{p,\delta^k,i} \\ N \leq i \leq N + \lceil \frac{1}{\delta} \rceil}} \Phi(X^i). \quad (4)$$

Equivalently, conditioned on any sequence in $R_{p,\delta^k,i}$ with i between N and $N + M - 1$, the scheme Φ fails with probability (under q) at least δ in step $i + 1$.

A sequence on which Φ has entered, but such that Φ has not been bankrupted on any of the sequence's prefixes is called a *surviving* sequence.

Consider a surviving sequence $x^N \in R_{p,\delta^k,N}$ in the support of p at level N . Given x^N , let the conditional probability that Φ is bankrupted in the following step be δ_N . From (4), as mentioned before, we have $\delta_N \geq \delta$.

Now, given $x^N \in R_{p,\delta^k,N}$, the conditional probability that Φ is bankrupted in at most two further steps is,

$$\delta_N + (1 - \delta_N)\delta_{N+1} \geq \delta + (1 - \delta)\delta,$$

where δ_{N+1} is interpreted as the weighted average (over surviving length- $(N + 1)$ suffixes of x^N) of the probability that Φ goes bankrupt in step $N + 2$.

Similarly, given a sequence $x^N \in R_{p,\delta^k,N}$, the probability that Φ is bankrupted on suffixes of x^N with length between N and $N + M$ is

$$\delta_N + (1 - \delta_N)\delta_{N+1} + \dots + \delta_{N+M} \prod_{i=N}^{N+M-1} (1 - \delta_i)$$

for some $\delta_N, \delta_{N+1}, \dots, \delta_{N+M}$, all of which are $\geq \delta$.

Let q_1 be the probability of all survivors in $R_{p,\delta^k,N}$, and q_2 be the probability of all sequences in $R_{p,\delta^k,N}$ where Φ has already been bankrupted. Therefore $q_1 + q_2 = q(R_{p,\delta^k,N})$.

Let $\bar{\delta}$ stand for $1 - \delta$. Now Φ is bankrupted with probability

$$\begin{aligned} &\geq q_2 + q_1 \left(\delta_N + \dots + \delta_{N+M} \prod_{i=N}^{N+M-1} (1 - \delta_i) \right) \\ &= q_2 + q_1 (\delta_N + \bar{\delta}_N (\delta_{N+1} + \bar{\delta}_{N+1} (\dots (\delta_{N+M-1} + \bar{\delta}_{N+M-1} \delta_{N+M})))) \\ &\geq q_2 + q_1 (\delta + (1 - \delta)\delta + \dots + (1 - \delta)^M \delta) \\ &= q_2 + q_1 (1 - (1 - \delta)^{\lceil 1/\delta \rceil}) \\ &\geq q(R_{p,\delta^k,N}) (1 - (1 - \delta)^{\lceil 1/\delta \rceil}) \\ &\geq \left(1 - \frac{1}{N} - h(\alpha) \right) (1 - (1 - \delta)^{\lceil 1/\delta \rceil}). \end{aligned}$$

The Theorem follows. \square

5. Sufficient condition for insurability

If no $p \in \mathcal{P}$ is deceptive, there is for each $p \in \mathcal{P}$ a number $\epsilon_p > 0$ such that, for every $\delta > 0$, over the set of probability distributions in \mathcal{P} that are in the neighborhood

$$\{p' \in \mathcal{P} : \mathcal{J}(p', p) < \epsilon_p\},$$

there is a uniform bound on the δ -percentile. We pick such an ϵ_p for each $p \in \mathcal{P}$ and call it the *reach* of p . For $p \in \mathcal{P}$, the set

$$B_p = \{p' \in \mathcal{P} : \mathcal{J}(p, p') < \epsilon_p\},$$

where ϵ_p is the reach of p , will play the role of the set of probability distributions in \mathcal{P} such that even if the true marginal loss distribution in force is one of these distributions, it will be okay to eventually set loss-dominants as if p were in force.

To prove that \mathcal{P}^∞ is insurable if there are no deceptive distribution among its one dimensional marginals, \mathcal{P} , we will need to find a way to cover the set \mathcal{P} with countable many sets of the form B_p . Unfortunately, $\mathcal{J}(p, q)$ is not a metric, so it is not immediately clear how to go about doing this. On the other hand since $\mathcal{J}(p', p) \leq |p - p'|_1 / \ln 2$, where $|p - p'|_1$ denotes the ℓ_1 distance between p and p' , see Lemma 5 in the Appendix, we can bootstrap off an understanding of the topology induced on \mathcal{P} by the ℓ_1 topology on the set of all probability distributions on \mathbb{N} .

5.1. Topology of \mathcal{P} with the ℓ_1 metric

The topology induced on \mathcal{P} by the ℓ_1 metric is Lindelöf, i.e. any covering of \mathcal{P} with open sets in the ℓ_1 topology has a countable subcover, see [5, Defn. 6.4] for this definition of a Lindelöf topological space.

We can show that \mathcal{P} with the ℓ_1 topology is Lindelöf by appealing to the fact that the set of all probability distributions on \mathbb{N} , with the ℓ_1 topology, is second countable, i.e. that it has a countable basis, which is a consequence of its have a countable norm-dense set (consider the set of all probability distributions on \mathbb{N} with finite support and with all probabilities being rational). Now, \mathcal{P} , as a topological subspace of a second countable topological space is also second countable [5, Theorem 6.2(2)], and every second countable topological space is Lindelöf [5, Thm. 6.3].

5.2. Sufficient condition

We now have the machinery required to prove that if no $p \in \mathcal{P}$ is deceptive, then \mathcal{P}^∞ is insurable, which is the other direction of Theorem 1, as stated next.

Theorem 3. If no $p \in \mathcal{P}$ is deceptive, then \mathcal{P}^∞ is insurable.

Proof The proof is constructive. For any $0 < \eta < 1$, we obtain a loss domination scheme Φ such that for all $p \in \mathcal{P}^\infty$, $p(\Phi \text{ goes bankrupt}) < \eta$.

For $p \in \mathcal{P}$, let

$$Q_p = \left\{ q : |p - q|_1 < \frac{\epsilon_p^2 (\ln 2)^2}{16} \right\},$$

where ϵ_p is the reach of p . The set Q_p is non-empty when $\epsilon_p > 0$. For large n , loss sequences of length n with empirical distribution in Q_p will play the role of those loss sequences on which the loss domination scheme Φ to be proposed will have entered; this will ensure that Φ enters with probability 1 when p is in force. Note that if ϵ_p is small enough then $Q_p \cap \mathcal{P} \subset B_p$.

Since no $p \in \mathcal{P}$ is deceptive, none of the sets Q_p are empty and the space \mathcal{P} of distributions can be covered by the sets $Q_p \cap \mathcal{P}$, namely

$$\mathcal{P} = \cup_{p \in \mathcal{P}} (Q_p \cap \mathcal{P}).$$

From Section 5.1, we know that \mathcal{P} is Lindelöf under the ℓ_1 topology. Thus, there is a countable set $\tilde{\mathcal{P}} \subset \mathcal{P}$, such that \mathcal{P} is covered by the collection of relatively open sets $\mathcal{Q}_{\tilde{p}}$, where

$$\mathcal{Q}_{\tilde{p}} := \{Q_{\tilde{p}} \cap \mathcal{P} : \tilde{p} \in \tilde{\mathcal{P}}\}.$$

We index the countable set $\tilde{\mathcal{P}}$ (or $\mathcal{Q}_{\tilde{p}}$) by $\iota : \tilde{\mathcal{P}} \rightarrow \mathbb{N}$.

We now describe the loss domination scheme Φ having the property that for all $p \in \mathcal{P}^\infty$, $p(\Phi \text{ goes bankrupt}) < \eta$.

Preliminaries Consider a length- n sequence x^n on which Φ has not entered thus far. Let the empirical distribution of the sequence be q , and let

$$\mathcal{P}'_q := \{p' \in \tilde{\mathcal{P}} : q \in Q_{p'}\}$$

be the set of distributions in \mathcal{P} (more precisely, $\tilde{\mathcal{P}}$) which can potentially capture q . Note that q in general need not belong to \mathcal{P} .

If $\mathcal{P}'_q \neq \emptyset$, we will refine the set of distributions that could capture q further to $\mathcal{P}_q \subset \mathcal{P}'_q$. This is to ensure that models in \mathcal{P}'_q do not prematurely capture loss sequences.

Let p be the model in force. The idea is that we want sequences generated by p to be captured by models that have p in their reach. We will require (5) below to hold to ensure that if any other distribution $p' \in \mathcal{P}_q$ which does not have p in its reach captures sequences with empirical distribution q , then the probability of such sequences is not too large under p . In addition, we impose (6) as well to resolve a technical issue since q need not, in general, belong to \mathcal{P} .

For $p' \in \mathcal{P}'_q$, let the reach of p' be $\epsilon_{p'}$, and define

$$D_{p'} := \frac{\tilde{\epsilon}^4 (\ln 2)^4}{256}.$$

This quantity will lower bound the distance of the empirical distribution q in question from the distribution p in force if p happens to be out of the reach of p' . Specifically, we place $p' \in \mathcal{P}_q$, if n satisfies

$$\exp\left(-n D_{p'}/18\right) \leq \frac{\eta}{2C(p')\iota(p')^2 n^2 \frac{36}{\pi^4}}, \quad (5)$$

and

$$F_q^{-1}(1 - \sqrt{D_{p'}/3}) \leq \log C(p'), \quad (6)$$

where $C(p')$ is

$$C(p') := 2^{\left(\sup_{r \in B_{p'}} F_r^{-1}(1 - \sqrt{D_{p'}/3}) \right)}.$$

Note that $C(p')$ is finite since p' is not deceptive. Comparison with Lemma 7 will give a hint as to why the equations above look the way they do.

Description of Φ For the sequence x^n with type q , if $\mathcal{P}_q = \emptyset$, the scheme does not enter yet. If $\mathcal{P}_q \neq \emptyset$, let p_q denote the distribution in \mathcal{P}_q with the smallest index.

All sequences with prefix x^n (namely sequences obtained by concatenating x^n with by any other sequence of symbols) are then said to be *trapped* by p_q —namely, loss-dominants will be based on p_q . The loss-dominant assigned for a length- m sequence trapped by p_q is

$$2f_{p_q}\left(\frac{6\eta/2}{2\pi^2 n^2}\right).$$

Φ enters with probability 1 First, we verify that the scheme enters with probability 1, no matter what distribution $p \in \mathcal{P}$ is in force. Every distribution $p \in \mathcal{P}$ is contained in at least one of the sets in $\mathcal{Q}_{\tilde{\mathcal{P}}}$. Let $Q \in \mathcal{Q}_{\tilde{\mathcal{P}}}$ be the set with the smallest index among all sets in $\mathcal{Q}_{\tilde{\mathcal{P}}}$ that contain p . There is thus some $\gamma > 0$ such that the neighborhood around p given by

$$I(p, \gamma) := \{q : |p - q|_1 < \gamma\}$$

satisfies $I(p, \gamma) \subseteq Q$. Let p' be the distribution of $\tilde{\mathcal{P}}$ which defines the set Q in $\mathcal{Q}_{\tilde{\mathcal{P}}}$. Note in particular that p is in the reach of p' .

With probability 1, the empirical distribution of sequences generated by p lies within $I(p, \gamma)$ [6] (see also Lemma 7 for an alternate proof). Now (5) will hold for all empirical distributions that fall in $I(p, \gamma)$, if we make n large enough—since $C(p')$ and $\iota(p')$ do not change with n and the right hand side diminishes to zero polynomially with n , while the left hand side diminishes exponentially to zero. Lastly, (6) will also hold almost surely, for q is the empirical probability of sequences generated by p , then $F_q^{-1}(1 - \sqrt{D_{p'}/3}) \rightarrow F_p^{-1}(1 - \sqrt{D_{p'}/3})$ with probability 1, and $F_p^{-1}(1 - \sqrt{D_{p'}/3}) \leq \log C(p')$ since p is within the reach of p' .

Thus the scheme enters with probability 1.

Probability of bankruptcy $\leq \eta$ We now analyze the scheme. Consider any $p \in \mathcal{P}$. Among sequences on which Φ has entered, we will distinguish between those that are in *good* traps and those in *bad* traps. If a sequence x^n is trapped by p' such that $p \in B_{p'}$, p' is a good trap. Conversely, if $p \notin B_{p'}$, p' is a bad trap.

(*Good traps*) Suppose a length- n sequence x^n is in a good trap, namely, it is trapped by a distribution p' such that $p \in B_{p'}$. Recall that the loss-dominant assigned is

$$2f_{p'}\left(\frac{6\eta}{2\pi^2 n^2}\right) \geq 2F_p^{-1}\left(1 - \frac{6\eta}{2\pi^2 n^2}\right),$$

where the inequality follows because p' is not deceptive, and p is within the reach of p' . Therefore, the scheme is bankrupted with probability at most $\delta' = 6\eta/2\pi^2 n^2$ in the next step. Therefore, summing over all n , sequences in good traps contribute at most $\eta/2$ to the probability of bankruptcy.

(*Bad traps*) We will show that the probability with which sequences generated by p fall into bad traps $\leq \eta/2$. Pessimistically, the conditional probability of bankruptcy in the very next step given a sequence falls into a bad trap is going to be upper bounded by 1. Thus the contribution to bankruptcy by sequences in bad traps is at most $\eta/2$.

Let q be any length- n empirical distribution trapped by \tilde{p} with reach $\tilde{\epsilon}$ such that $p \notin B_{\tilde{p}}$, we obtain from Lemma 6 that $\mathcal{J}(p, q) \geq \frac{\tilde{\epsilon}^2 \ln 2}{16}$. Hence, for all q trapped by \tilde{p} ,

$$|p - q|_1^2 \geq \mathcal{J}^2(p, q)(\ln 2)^2 \geq \frac{\tilde{\epsilon}^4 (\ln 2)^4}{256} = D_{\tilde{p}}^2$$

Thus, for $p \in \mathcal{P}^\infty$, the probability length- n sequences with empirical distribution q is trapped by a bad \tilde{p} is, using (5) and (6)

$$\begin{aligned} &\leq p\left(|q - p|^2 \geq D_{\tilde{p}} \text{ and } F_q^{-1}\left(1 - \frac{\sqrt{D_{\tilde{p}}}}{3}\right) \leq \log C(\tilde{p})\right) \\ &\stackrel{(a)}{\leq} (C(\tilde{p}) - 2) \exp\left(-\frac{nD_{\tilde{p}}}{18}\right) \\ &\stackrel{(b)}{\leq} \frac{\eta(C(\tilde{p}) - 2)}{2C(\tilde{p})\iota(\tilde{p})^2 n^2} \frac{36}{\pi^4} \\ &\leq \frac{\eta}{2\iota(\tilde{p})^2 n^2} \frac{36}{\pi^4}, \end{aligned}$$

where the inequality (a) follows from Lemma 7 and (b) from (5). Therefore, the probability of sequences falling into bad traps

$$\leq \sum_{n \geq 1} \sum_{p' \in \mathcal{Q}_{\tilde{p}}} \frac{\eta}{2\iota(p')^2 n^2} \frac{36}{\pi^4} \leq \eta/2$$

since $\sum_{p' \in \mathcal{Q}_{\tilde{p}}} \frac{1}{\iota(p')^2} \leq \sum_{n \geq 1} \frac{1}{n^2} = \frac{\pi^2}{6}$. The theorem follows. \square

6. Concluding remarks

The loss domination problem formulated and solved in this paper appears to be of natural interest. However, there are several features of the insurance problem formulated here that might appear troubling even to the casual reader. In practice an insured party entering into an insurance contract would expect some stability in the premiums that are expected to be paid. A natural direction for further research is therefore to study how the notion of insurability of a model class changes when one imposes restrictions on how much the premium set by the insurer can vary from period to period. Another obvious shortcoming of the formulation of the insurance problem studied here is the assumption that the insured will accept any contract issued by the insurer. Since the insured in our model represents an aggregate of individual insured parties, a natural direction to make the framework more realistic would be to think of the insured parties as being of different *types*. This would in effect make the total realized premium from the insured (the aggregate of the insured parties) and the distribution of the realized loss in each period a function of the size of the premium per insured party set by the insurer in that period. Characterizing which model classes are insurable when the realized premium and the realized loss are functions of a set premium per insured party would be of considerable interest.

Both for the loss domination problem and for the insurance problem, working with model classes for the loss sequence that allow for dependencies in the loss from period to period, for instance Markovian dependencies, would be another interesting direction for further research. Considering models with multiple, possibly competing insurers, as well as considering an insurer operating in multiple markets, where losses in one market can be offset by gains in another, also seem to be useful directions to investigate.

Acknowledgments

The material of the paper has appeared in the conference [7]. N. Santhanam was supported by NSF Grants CCF-1065632, CCF-1018984 and EECS-

1029081. V. Anantharam was supported by the ARO MURI grant W911NF-08-1-0233, Tools for the Analysis and Design of Complex Multi-Scale Networks, the NSF grant CNS-0910702, the NSF Science & Technology Center grant CCF-0939370, Science of Information, Marvell Semiconductor Inc., and the U.C. Discovery program. We also thank C. Nair and K. Viswanathan for helpful discussions.

Appendix

Lemma 4. Let p and q be probability distributions on \mathbb{N} with support $\mathcal{A} \subset \mathbb{N}$. Suppose $\mathcal{J}(p, q) \leq \epsilon$. For any $S \subset \mathcal{A}$ and $\alpha < 1 - \ln 2 = .30685$, if $p(S) \geq 1 - \alpha$, then

$$q(S) \geq 1 - 2\epsilon - 2h(\alpha).$$

Proof Let p_S (respectively q_S) denote a binary distribution, the two probabilities of which correspond to $[p(S), 1 - p(S)]$ (respectively $[q(S), 1 - q(S)]$). Now,

$$\begin{aligned} \epsilon \geq \mathcal{J}(p, q) &\geq D\left(p_S \parallel \frac{q_S + p_S}{2}\right) \\ &\geq p(S) \log \frac{2}{p(S) + q(S)} - h(p_S) \\ &\geq (1 - \alpha) \log \frac{2}{p(S) + q(S)} - h(\alpha). \end{aligned}$$

The last inequality follows because the condition $p(S) \geq 1 - \alpha \geq \frac{1}{2}$ implies $h(p_S) \leq h(1 - \alpha)$. Therefore,

$$\log \frac{2}{p(S) + q(S)} \leq \frac{h(\alpha) + \epsilon}{(1 - \alpha)},$$

implying that

$$\begin{aligned} \frac{1 + q(S)}{2} &\geq \frac{p(S) + q(S)}{2} \geq 2^{-\frac{h(\alpha) + \epsilon}{1 - \alpha}} \\ &\geq 1 - (h(\alpha) + \epsilon), \end{aligned}$$

where the last inequality follows since $\ln 2 \leq 1 - \alpha$. □

Lemma 5. Let p and q be probability distributions on \mathbb{N} . Then

$$\frac{1}{4 \ln 2} |p - q|_1^2 \leq \mathcal{J}(p, q) \leq \frac{1}{\ln 2} |p - q|_1.$$

If, in addition, r is a probability distribution on \mathbb{N} , then

$$\mathcal{J}(p, q) + \mathcal{J}(q, r) \geq \mathcal{J}^2(p, r) \frac{\ln 2}{8}.$$

Proof The lower bound in the first statement follows since

$$D\left(p \parallel \frac{p+q}{2}\right) \geq \frac{1}{2 \ln 2} \frac{1}{4} |p - q|_1^2$$

and similarly for $D(q \parallel \frac{p+q}{2})$. The upper bound in the first statement follows since

$$\begin{aligned} \mathcal{J}(p, q) \ln 2 &\leq \sum_x p(x) \left(\frac{p(x) - q(x)}{p(x) + q(x)} \right) \\ &\quad + \sum_x q(x) \left(\frac{q(x) - p(x)}{p(x) + q(x)} \right) \\ &\leq |p - q|_1. \end{aligned}$$

To prove the triangle-like inequality, note that

$$\begin{aligned} \mathcal{J}(p, q) + \mathcal{J}(q, r) &\geq \frac{1}{4 \ln 2} (|p - q|_1^2 + |q - r|_1^2) \\ &\geq \frac{1}{8 \ln 2} (|p - q|_1 + |q - r|_1)^2 \\ &\geq \frac{1}{8 \ln 2} (|p - r|_1)^2 \\ &\geq \frac{\ln 2}{8} \mathcal{J}(p, r)^2, \end{aligned}$$

where the last inequality follows from the upper bound on $\mathcal{J}(p, r)$ already proved. \square

Lemma 6. Let $\epsilon_0 > 0$. If

$$|p_0 - q|_1 \leq \frac{\epsilon_0^2 (\ln 2)^2}{16},$$

then for all $p \in \mathcal{P}$ with $\mathcal{J}(p, p_0) \geq \epsilon_0$, we have

$$\mathcal{J}(p, q) \geq \frac{\epsilon_0^2 \ln 2}{16}.$$

Proof Since

$$|p_0 - q|_1 \leq \frac{\epsilon_0^2 (\ln 2)^2}{16},$$

Lemma 5 implies that

$$\mathcal{J}(p_0, q) \leq \frac{\epsilon_0^2 \ln 2}{16}.$$

Further, Lemma 5 then implies that

$$\mathcal{J}(p, q) + \frac{\epsilon_0^2 \ln 2}{16} \geq \mathcal{J}(p, q) + \mathcal{J}(p_0, q) \geq \frac{\mathcal{J}^2(p, p_0) \ln 2}{8} \geq \frac{\epsilon_0^2 \ln 2}{8},$$

where the last inequality follows since $\mathcal{J}(p, p_0) \geq \epsilon_0$. \square

Lemma 7. Let p be any probability distribution on \mathbb{N} . Let $\delta > 0$ and let $k \geq 2$ be an integer. Let X_1^n be a sequence generated *i.i.d.* with marginals p and let $q(X^n)$ be the empirical distribution of X_1^n . Then

$$\begin{aligned} p(|q(X^n) - p| > \delta \text{ and } F_q^{-1}(1 - \delta/3) \leq k) \\ \leq (2^k - 2) \exp\left(-\frac{n\delta^2}{18}\right). \end{aligned}$$

Proof There is a similar lemma in [8]. The difference from [8] is that the right side of the inequality above does *not* depend on p , and this property is crucial for its use here.

The starting point is the following result. Suppose p' is a probability distribution on \mathbb{N} with finite support of size L . Then from [9], if we consider length n sequences,

$$p'(|q(X^n) - p'|_1 \leq \delta) \geq 1 - (2^L - 2) \exp\left(-\frac{n\delta^2}{2}\right). \quad (7)$$

Since $k \geq 2$, consider the distributions p' and q' with support $A = \{1, \dots, k-1\} \cup \{-1\}$, obtained as

$$p'(i) = \begin{cases} p(i) & i < k \\ \sum_{j=k}^{\infty} p(j) & i = -1 \end{cases},$$

and similarly for q' .

From (7),

$$p'(|p' - q'|_1 > \delta/3) \leq (2^k - 2) \exp\left(-\frac{n\delta^2}{18}\right).$$

We will show that if $F_q^{-1}(1 - \delta/3) \leq k$ and $|p - q|_1 > \delta$ then $q'(-1) \leq \delta/3$ and $|p' - q'|_1 > \delta/3$. Thus, we will have

$$\begin{aligned} p(|q(X^n) - p|_1 > \delta \text{ and } F_q^{-1}(1 - \delta/3) \leq k) \\ \leq p'(|p' - q'|_1 > \delta/3 \text{ and } q'(-1) \leq \delta/3) \\ \leq (2^k - 2) \exp\left(-\frac{n\delta^2}{18}\right). \end{aligned}$$

Finally, as in [8],

$$\begin{aligned}
 |p - q|_1 &= \sum_{l=1}^{k-1} |p(l) - q(l)| \\
 &\leq \sum_{j=k}^{\infty} (p(j) - q(j)) + 2 \sum_{j=k}^{\infty} q(j) \\
 &\leq |p'(-1) - q'(-1)| + 2\delta/3.
 \end{aligned}$$

Since $p(l) = p'(l)$ and $q(l) = q'(l)$ for all $l = 1, \dots, k-1$, we have $|p' - q'|_1 \geq |p - q|_1 - 2\delta/3$. If $|p - q|_1 \geq \delta$ in addition, $|p' - q'|_1 \geq \delta/3$. \square

References

- [1] K. Englund and A. Martin-Löf. *Statisticians of the Centuries*, chapter Ernst Filip Oskar Lundberg, pages 308–311. New York: Springer, 2001.
- [2] H. Cramer. Historical Review of Filip Lundberg’s Work on Risk Theory. *Skandinavisk Aktuarietidskrift (Suppl.)*, 52:6–12, 1969. Reprinted in The Collected Works of Harald Cramér edited by Anders Martin-Löf, 2 volumes Springer 1994.
- [3] S. Asmussen and H. Albrecher. *Ruin probabilities*. World Scientific Publishing Company, 2nd edition, 2010.
- [4] N. Santhanam and V. Anantharam. Agnostic insurance tasks and their relation to compression. In *International conference on signal processing and communications (SPCOM)*, 2012.
- [5] J. Dugundji. *Topology*. Allyn and Bacon Inc., Boston, 1970.
- [6] K.L. Chung. A note on the ergodic theorem of information theory. *Annals of Mathematical Statistics*, 32:612–614, 1961.
- [7] N. Santhanam and V. Anantharam. Prediction over countable alphabets. In *Conference on Information Sciences and Systems*, 2012.
- [8] S. Ho and R. Yeung. On information divergence measures and joint typicality. *IEEE Transactions on Information Theory*, 56(12):5893–5905, 2010.
- [9] T. Weissman, E. Ordentlich, G. Seroussi, S. Verdú, and M. Weinberger. Universal discrete denoising: known channel. *IEEE Transactions on Information Theory*, 51(1):5–28, 2005. See also HP Labs Tech Report HPL-2003-29, Feb 2003.